

Bayesian hierarchical functional data analysis with basis function approximations using Gaussian-Wishart processes

Jingjing Yang

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA.

email: yjingj@umich.edu

and

Jong Soo Lee

Department of Mathematical Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA.

and

Peng Ren

Suntrust Bank, Atlanta, GA 30308, USA.

and

Dennis D. Cox

Department of Statistics, Rice University, Houston, TX 77005, USA.

and

Taeryon Choi

Department of Statistics, Korea University, Seoul 136-701, Republic of Korea.

email: trchoi@korea.ac.kr

SUMMARY: Functional data are defined as realizations of random functions varying over a continuum, which are usually collected on discretized grids but not necessarily on common grids with low dimension. To tackle the issue of uncommon observation grids and high dimensionality for Bayesian functional data analysis with Gaussian processes, we propose conducting posterior inferences with basis function approximations. Specifically, we illustrate our basis function approximation method for a Bayesian hierarchical model that simultaneously smooth functional observations and estimate the mean-covariance functions using Gaussian-Wishart processes. Assuming functional data are of Gaussian processes and approximated by a system of basis functions, we can conduct Bayesian inferences for the functional signals, mean, and covariance through the posterior samplings of the basis function

000 0000

coefficients. Both simulations and real studies demonstrate that, our approach with basis function approximations could conduct posterior inferences > 10 times more efficient on both computation speed and memory usage, providing stable and consistent posterior estimates. When the functional data have low dimensional common grids, our approach generates similar results as the usual Bayesian inferences. When the observation grids are high dimensional or random, our approach can still efficiently conduct inferences while the usual Bayesian inferences suffer serious computation burden and unstableness. Our basis function approximation method sheds lights on how to resolve the curse of dimensionality for Bayesian functional data analysis.

KEY WORDS: basis functions; Bayesian hierarchical model; functional data analysis; Gaussian process; Inverse-Wishart process; smoothing;

1. Introduction

Functional data — defined as realizations of random functions varying over a continuum (Ramsay and Silverman, 2005) — include a wide range of data types, e.g., longitudinal data that are functional observations over time, spectroscopic data that are represented as functional observations over a series of wavelengths, and meteorological data that are functional observations over both time and space. Many statistical methods have been developed for *functional data analysis* (FDA) since Ramsay and Dalzell (1991) first dived into this field. Representing functional data by basis functions has been widely used in FDA to resolve the issue of high dimensionality, e.g., transforming discretely observed data into smooth functions (Ramsay and Silverman, 2005), functional linear regression models (Cardot et al., 1999, 2003; Hall et al., 2007; Zhu and Cox, 2009; Zhu et al., 2011), functional principle components analysis (Crainiceanu and Goldsmith, 2010; Zhu et al., 2014), and functional data classification (Abraham et al., 2003; Garcia-Escudero and Gordaliza, 2005; Zhu et al., 2010; Serban and Wasserman, 2012). However, this technique of basis function representation has not been adapted to FDA with Gaussian processes (GPs), especially nonparametric Bayesian analysis with GPs (Shi et al., 2007; Banerjee et al., 2008; Yang et al., 2015).

As is known, Bayesian analysis methods usually suffer heavier computation burden than the frequentist methods because of sampling posterior distributions by Markov chain Monte Carlo (MCMC). This is a critical issue for analyzing functional data, whose dimensionality is essentially infinite. Take Bayesian GP regression models for an example, the computation cost for posterior inferences with MCMC is of order $O(np^3m)$, with sample size n , data dimension p , and MCMC iteration number m . Various methods have been proposed to address the computational bottleneck issue in GP regression models: (i) Approximate a large matrix by a low rank matrix in computations (Quiñonero Candela et al., 2007; Shi and Choi, 2011; Rasmussen and Williams, 2006, Chapter 8). (ii) Reduce computation cost for covariance matrix estimation, e.g., Cressie and Johannesson

(2008) considered fixed rank kriging, Banerjee et al. (2008) proposed a predictive process conditioning on a finite number of observation points, and Banerjee et al. (2013) used a linear random projection. However, good low-rank approximation matrices are not easily obtained; the methods by Cressie and Johannesson (2008); Banerjee et al. (2008) underestimate variance by only using a subset of the observation data; while the method by Banerjee et al. (2013) requires extra computation burden for finding a projection matrix and has computation cost of order $O(n^2p)$ (the computation cost will still be an issue with extremely large p).

Besides the computation burden related to covariance matrix, Bayesian analysis with GPs can also encounter the issues of failing MCMC with singular covariance samples and large memory usage for saving MCMC samples. Here, we propose a novel approach to conduct posterior inferences with basis function approximations for Bayesian GP regression models, which will not only reduce the computation cost to the order of $O(nK^3M)$ with $K \ll p$ but also stably produce consistent estimates.

Specifically, we illustrate our method for the Bayesian hierarchical model with Gaussian-Wishart processes (BHM) proposed by Yang et al. (2015), which simultaneously and nonparametrically smooths functional data and estimates mean-covariance functions. Assuming functional data $Z_i(\tau)$ are represented by $Z_i(\tau) = \mathbf{B}(\tau)\zeta_i$, with basis functions $\mathbf{B}(\tau) = (b_1(\tau), \dots, b_K(\tau))$ and an arbitrary low dimensional working grid τ , a Bayesian framework for the coefficients ζ_i can be derived from the model assumptions for $Z_i(\tau)$. Then we can conduct posterior inferences for functional signals, mean, and covariance; through MCMC samplings with ζ_i , mean and covariance of ζ_i . In this paper, we show details of this method with cubic B-splines and an equally spaced working grid τ , which can be further generalized to other basis functions and Bayesian regression models with GPs.

Simulation studies with both stationary and nonstationary functional data show that: When functional data are observed on low dimensional common grids, the BHM with basis function

approximations generates similar results as the previous BHM method by Yang et al. (2015). When functional data are observed on high dimensional grids or random grids, the previous BHM method is likely to fail because of computational issues, while the BHM with basis function approximations performs efficiently in computation and produces estimates with smaller root mean square errors (RMSEs) than the alternative frequentist method — Principle Analysis by Conditional Expectation (PACE) proposed by Yao et al. (2005b). A real case study with sleeping energy expenditure measurements of 106 children and adolescents (44 obese cases, 62 controls) over 405 time points (Lee et al., 2016) show that the BHM with basis function approximations captures the periodic pattern of the measurements, hence gives more reasonable estimates for the functional signals, mean, and covariance.

This paper is organized as follows: We describe our approach of basis function approximations for BHM in Section 2, and posterior inferences with MCMC in Section 3. Simulation and real case studies are presented in Section 4 and Section 5 respectively. Discussion is provided in Section 6.

2. BHM with basis function approximations

Let $\{Y_i(\cdot); i = 1, 2, \dots, n\}$ denote the observed functional data consisting of n observations.

Consider the general functional measurement error model

$$Y_i(t_{ij}) = Z_i(t_{ij}) + \epsilon_{ij}; \quad t_{ij} \in \mathcal{T}; \quad i = 1, \dots, n; \quad j = 1, \dots, p_i; \quad (1)$$

where $\{Z_i(\mathbf{t}_i)\}$ are the underlying true function data, and $\{\mathbf{t}_i = (t_{i1}, \dots, t_{ip_i})^T\}$ are the grids on which functions $\{Z_i(\cdot)\}$ are evaluated.

2.1 BHM

With the goal of smoothing functional data and estimating mean-covariance functions, the BHM (Yang et al., 2015) assumes

$$\begin{aligned} Z_i(\cdot) \text{ i.i.d. } &\sim GP(\mu_Z(\cdot), \Sigma_Z(\cdot, \cdot)); \mu_Z(\cdot) | \Sigma_Z(\cdot, \cdot) \sim GP(\mu_0(\cdot), \frac{1}{c} \Sigma_Z(\cdot, \cdot)); \\ \epsilon(t_{ij}) &\sim N(0, \sigma_\epsilon^2); \sigma_\epsilon^2 \sim \text{InverseGamma}(a_\epsilon, b_\epsilon); \\ \Sigma_Z(\cdot, \cdot) &\sim IWP(\delta, \Psi(\cdot, \cdot)); \Psi(\cdot, \cdot) = \sigma_s^2 A(\cdot, \cdot); \sigma_s^2 \sim \text{Gamma}(a_s, b_s); \end{aligned} \quad (2)$$

- $GP(\mu_Z(\cdot), \Sigma_Z(\cdot, \cdot))$ denotes a GP with mean function $\mu_Z(\cdot)$ and covariance function $\Sigma_Z(\cdot, \cdot)$;
- Prior mean $\mu_0(\cdot)$ is selected as a smooth function, e.g., smoothed sample mean;
- c is taken as a constant, e.g. $c = n$ or $c = 1$;
- $IWP(\delta, \Psi(\cdot, \cdot))$ denotes a Inverse-Wishart process (IWP) with shape parameter δ and scaling parameter $\Psi(\cdot, \cdot)$, following the parameterization by Dawid (1981);
- $A(\cdot, \cdot)$ is a smooth correlation or covariance function, e.g., Matérn function (Matérn, 1960).

Yang et al. (2015) conducts posterior inferences for BHM in the functional data space, evaluating functions on the pooled-grid of all observation grids $\{\mathbf{t}_i\}$. Thus, the computation cost is of order $O(np^3m)$, with sample size n , pooled-grid dimension p , and total MCMC iterations m . When p is large, besides the computation burden, computational issues such as failing MCMC, unstable estimation, and exceeding memory capacity are likely to appear.

2.2 Approximation by basis functions

Here, we propose a novel approach for conducting Bayesian inferences in BHM with approximations by basis functions (BABF), which provides one way to resolve the computational bottleneck for Bayesian GP regression models.

First, we choose an arbitrary working grid, $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_L)^T \subset \mathcal{T}$, whose dimension L can be much smaller than the pooled-grid dimension. Without loss of generality, we choose an equally spaced grid on the domain \mathcal{T} for computational convenience. Then we can approximate the GP

evaluations $Z_i(\boldsymbol{\tau})$ by a system of basis functions (e.g., cubic B-splines), such that

$$Z_i(\boldsymbol{\tau}) = \sum_{k=1}^K \zeta_{ik} b_k(\boldsymbol{\tau}) = \mathbf{B}(\boldsymbol{\tau}) \boldsymbol{\zeta}_i; \quad (3)$$

$$\mathbf{B}(\boldsymbol{\tau}) = [b_1(\boldsymbol{\tau}), b_2(\boldsymbol{\tau}), \dots, b_K(\boldsymbol{\tau})]; \quad \boldsymbol{\zeta}_i = (\zeta_{i1}, \zeta_{i2}, \dots, \zeta_{iK})^T;$$

where $\{b_1(\cdot), b_2(\cdot), \dots, b_K(\cdot)\}$ are K basis functions, and $\boldsymbol{\zeta}_i$ is the coefficient vector. Assuming $K = L$, we can write $\boldsymbol{\zeta}_i = \mathbf{B}(\boldsymbol{\tau})^{-1} Z_i(\boldsymbol{\tau})$, a linear transformation of $Z_i(\boldsymbol{\tau})$. Note that even if $\mathbf{B}(\boldsymbol{\tau})$ is not invertible, $\boldsymbol{\zeta}_i$ can be still written as a linear transformation of $Z_i(\boldsymbol{\tau})$ with the generalized inverse (James, 1978) of $\mathbf{B}(\boldsymbol{\tau})$. Given $\boldsymbol{\zeta}_i$, the signal values on observation grid \mathbf{t}_i can be approximated by $Z_i(\mathbf{t}_i) \doteq \mathbf{B}(\mathbf{t}_i) \boldsymbol{\zeta}_i$.

Second, under the GP assumption in (2), $Z_i(\boldsymbol{\tau})$ follows a multivariate normal (MN) distribution, $MN(\mu_Z(\boldsymbol{\tau}), \Sigma_Z(\boldsymbol{\tau}, \boldsymbol{\tau}))$. Because $\boldsymbol{\zeta}_i$ is a linear transformation of $Z_i(\boldsymbol{\tau})$, the following model is induced for $\boldsymbol{\zeta}_i$,

$$\boldsymbol{\zeta}_i \sim MN(\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta); \quad \boldsymbol{\mu}_\zeta = \mathbf{B}(\boldsymbol{\tau})^{-1} \mu_Z(\boldsymbol{\tau}); \quad \boldsymbol{\Sigma}_\zeta = \mathbf{B}(\boldsymbol{\tau})^{-1} \Sigma_Z(\boldsymbol{\tau}, \boldsymbol{\tau}) \mathbf{B}(\boldsymbol{\tau})^{-T}. \quad (4)$$

Further, from the assumed priors for $(\mu_Z(\cdot), \Sigma_Z(\cdot, \cdot))$ in (2), the following priors for $(\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta)$ are also induced:

$$\boldsymbol{\mu}_\zeta | \boldsymbol{\Sigma}_\zeta \sim MN(\mathbf{B}(\boldsymbol{\tau})^{-1} \mu_0(\boldsymbol{\tau}), c \boldsymbol{\Sigma}_\zeta); \quad (5)$$

$$\boldsymbol{\Sigma}_\zeta \sim IW(\delta, \mathbf{B}(\boldsymbol{\tau})^{-1} \Psi(\boldsymbol{\tau}, \boldsymbol{\tau}) \mathbf{B}(\boldsymbol{\tau})^{-T}). \quad (6)$$

Third, for $i = 1, \dots, n$ in each MCMC iteration, conditioning on $Y_i(\mathbf{t}_i)$, $\mu_Z(\mathbf{t}_i)$, $\Sigma_Z((\mathbf{t}_i, \boldsymbol{\tau}), (\mathbf{t}_i, \boldsymbol{\tau}))$, we can sample $\boldsymbol{\zeta}_i$ from the its full conditional distribution; update $(\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta)$ conditioning on the present samples of $\{\boldsymbol{\zeta}_i; i = 1, \dots, n\}$; approximate functional values of $Z_i(\cdot)$, $\mu_Z(\cdot)$, and $\Sigma_Z(\cdot, \cdot)$ on the observation grids $\{\mathbf{t}_i\}$, using their evaluations on $\boldsymbol{\tau}$, basis functions, and present samples of $(\boldsymbol{\zeta}_i, \boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta)$; then update σ_ϵ^2 conditioning on $\{Z_i(\mathbf{t}_i); i = 1, \dots, n\}$ and update σ_s^2 conditioning on $\Sigma_Z(\boldsymbol{\tau}, \boldsymbol{\tau})$. Essentially, this MCMC process is still a Gibbs-Sampler (Geman and Geman, 1984), whose computation cost is of order $O(nK^3m)$ (details of the MCMC process are described in Section 3).

Last, We take the corresponding averages of the posterior samples as our Bayesian estimates, whose uncertainties can be easily quantified by MCMC credible intervals.

2.3 Choose hyper priors

The same strategies of setting hyper parameter values by Yang et al. (2015) are used. Specifically, $A(\cdot, \cdot)$ in (2) is taken as a Matérn covariance function if the studying functional data are stationary; while taken as a smooth covariance estimate of the observations, with non-stationary data — such a smooth covariance estimate can be obtained from existing methods such as PACE or sample estimate using pre-smoothed functional data. The hyper parameter c in (2) is taken as 1 to allow more uncertainty for the mean function. A heuristic Bayesian approach is taken for setting the values of $(a_\epsilon, b_\epsilon, a_s, b_s)$ in (2).

3. Posterior inferences by MCMC

The BHM assumptions in (2) lead to the following joint posterior distribution

$$f(\mathbf{Z}, \mu_Z, \Sigma_Z, \sigma_\epsilon^2, \sigma_s^2 | \mathbf{Y}) \propto f(\mathbf{Y} | \mathbf{Z}, \sigma_\epsilon^2) f(\mathbf{Z} | \mu_Z, \Sigma_Z) f(\mu_Z | \Sigma_Z) f(\Sigma_Z | \sigma_s^2) f(\sigma_\epsilon^2) f(\sigma_s^2), \quad (7)$$

where $\mathbf{Z} = \{Z_1(\mathbf{t}_i), \dots, Z_n(\mathbf{t}_n)\}$, $\mathbf{Y} = \{Y_1(\mathbf{t}_i), \dots, Y_n(\mathbf{t}_n)\}$. Because of the linear transformation relationship between ζ_i and $Z_i(\tau)$, $\zeta_i = \mathbf{B}(\tau)^{-1} Z_i(\tau)$, the joint distribution (7) is equivalent to

$$f(\zeta, \mu_\zeta, \Sigma_\zeta, \sigma_\epsilon^2, \sigma_s^2 | \mathbf{Y}) \propto f(\mathbf{Y} | \zeta, \sigma_\epsilon^2) f(\zeta | \mu_\zeta, \Sigma_\zeta) f(\mu_\zeta | \Sigma_\zeta) f(\Sigma_\zeta | \sigma_s^2) f(\sigma_\epsilon^2) f(\sigma_s^2), \quad (8)$$

where $\zeta = \{\zeta_1, \dots, \zeta_n\}$, $\mu_\zeta = \mathbf{B}(\tau)^{-1} \mu_Z(\tau)$, and $\Sigma_\zeta = \mathbf{B}(\tau)^{-1} \Sigma_Z(\tau, \tau) \mathbf{B}(\tau)^{-T}$.

3.1 Full conditional distribution of ζ_i

From (8), we can see $f(\zeta|\mathbf{Y}, \boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta) \propto f(\mathbf{Y}|\zeta, \sigma_\epsilon^2)f(\zeta|\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta)$, then the full conditional posterior distribution of ζ_i can be derived as

$$\begin{aligned} \zeta_i|(\mathbf{Y}_i(\mathbf{t}_i), \boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta) &\sim MN[\mathbf{m}_{\zeta_i|Y_i}, \mathbf{V}_{\zeta_i|Y_i}]; \\ \mathbf{V}_{\zeta_i|Y_i} &= \left(\frac{\mathbf{B}(\mathbf{t}_i)^T \mathbf{B}(\mathbf{t}_i)}{\sigma_\epsilon^2} + \boldsymbol{\Sigma}_\zeta^{-1} \right)^{-1}, \quad \mathbf{m}_{\zeta_i|Y_i} = \mathbf{V}_{\zeta_i|Y_i} \left(\frac{\mathbf{B}(\mathbf{t}_i)^T \mathbf{Y}_i(\mathbf{t}_i)}{\sigma_\epsilon^2} + \boldsymbol{\Sigma}_\zeta^{-1} \boldsymbol{\mu}_\zeta \right). \end{aligned} \quad (9)$$

3.2 Conditional distribution for $\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta$

Conditioning on $\{\zeta_i\}$, the posterior distributions for $\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta$ can be derived from

$$f(\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta | \zeta_1, \dots, \zeta_n) \propto \prod_{i=1}^n f(\zeta_i | \boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta) f(\boldsymbol{\mu}_\zeta | \boldsymbol{\Sigma}_\zeta) f(\boldsymbol{\Sigma}_\zeta), \quad (10)$$

where prior distributions $f(\boldsymbol{\mu}_\zeta | \boldsymbol{\Sigma}_\zeta)$, $f(\boldsymbol{\Sigma}_\zeta)$ are given by (5), (6). As a result,

$$\boldsymbol{\mu}_\zeta | (\zeta_1, \dots, \zeta_n, \boldsymbol{\Sigma}_\zeta) \sim MN \left(\frac{1}{n+c} (\sum_{i=1}^n \zeta_i + c\mathbf{B}(\boldsymbol{\tau})^{-1} \mu_0(\boldsymbol{\tau})), \frac{1}{n+c} \boldsymbol{\Sigma}_\zeta \right); \quad (11)$$

$$\boldsymbol{\Sigma}_\zeta | (\zeta_1, \dots, \zeta_n, \boldsymbol{\mu}_\zeta) \sim IW(\tilde{\delta}_\zeta, \tilde{\boldsymbol{\Psi}}_\zeta), \quad (12)$$

$$\begin{aligned} \tilde{\delta}_\zeta &= n + 1 + \delta, \quad \tilde{\boldsymbol{\Psi}}_\zeta = \sum_{i=1}^n (\zeta_i - \boldsymbol{\mu}_\zeta)(\zeta_i - \boldsymbol{\mu}_\zeta)^T + \\ &c(\boldsymbol{\mu}_\zeta - \mathbf{B}(\boldsymbol{\tau})^{-1} \mu_0(\boldsymbol{\tau}))(\boldsymbol{\mu}_\zeta - \mathbf{B}(\boldsymbol{\tau})^{-1} \mu_0(\boldsymbol{\tau}))^T + \mathbf{B}(\boldsymbol{\tau})^{-1} \boldsymbol{\Psi}(\boldsymbol{\tau}, \boldsymbol{\tau}) \mathbf{B}(\boldsymbol{\tau})^{-T}. \end{aligned}$$

3.3 MCMC process

We design the following MCMC process to sample posterior distributions, which is essentially a Gibbs-Sampler that has computation conveniences and ensures the property of posterior convergence.

Step 0: Choose hyper parameters (Section 2.3) and set initial values for all of the parameters in the model. Initial values for $(\mu_Z(\boldsymbol{\tau}), \Sigma_Z(\boldsymbol{\tau}, \boldsymbol{\tau}), \sigma_\epsilon^2)$ can be taken as empirical estimates; then initial values for $(\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta)$ are induced by (4).

Step 1: Conditioning on observed data \mathbf{Y} and current values of $(\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta, \sigma_\epsilon^2)$, sample $\{\zeta_i\}$ from (9).

Step 2: Conditioning on current values of ζ , update $\boldsymbol{\mu}_\zeta$ and $\boldsymbol{\Sigma}_\zeta$ respectively from (11) and (12).

Step 3: Given current values of $(\{\zeta_i\}, \mu_\zeta, \Sigma_\zeta)$, for $i = (1, \dots, n)$, approximate $Z_i(\mathbf{t}_i)$, $\mu_Z(\mathbf{t}_i)$, $\Sigma_Z(\mathbf{t}_i, \mathbf{t}_i)$, $\Sigma(\boldsymbol{\tau}, \mathbf{t}_i)$, $\Sigma_Z(\mathbf{t}_i, \boldsymbol{\tau})$, and $\Sigma_Z(\boldsymbol{\tau}, \boldsymbol{\tau})$ by

$$Z_i(\mathbf{t}_i) = \mathbf{B}(\mathbf{t}_i)\zeta_i, \mu_Z(\mathbf{t}_i) = \mathbf{B}(\mathbf{t}_i)\mu_\zeta, \Sigma(\mathbf{t}_i, \mathbf{t}_i) = \mathbf{B}(\mathbf{t}_i)\Sigma_\zeta\mathbf{B}(\mathbf{t}_i)^T,$$

$$\Sigma(\boldsymbol{\tau}, \mathbf{t}_i)^T = \Sigma(\mathbf{t}_i, \boldsymbol{\tau}) = \mathbf{B}(\mathbf{t}_i)\Sigma_\zeta\mathbf{B}(\boldsymbol{\tau})^T, \Sigma(\boldsymbol{\tau}, \boldsymbol{\tau}) = \mathbf{B}(\boldsymbol{\tau})\Sigma_\zeta\mathbf{B}(\boldsymbol{\tau})^T.$$

Step 4: Conditioning on \mathbf{Z} and \mathbf{Y} , update σ_ϵ^2 from

$$IG\left(a_\epsilon + \frac{1}{2} \sum_{i=1}^n p_i, b_\epsilon + \frac{1}{2} \sum_{i=1}^n (Y_i(\mathbf{t}_i) - Z_i(\mathbf{t}_i))^T (Y_i(\mathbf{t}_i) - Z_i(\mathbf{t}_i))\right),$$

which is derived from

$$f(\sigma_\epsilon^2 | Y_1(\mathbf{t}_1), Z_1(\mathbf{t}_1), \dots, Y_n(\mathbf{t}_n), Z_n(\mathbf{t}_n)) \propto \prod_{i=1}^n f(Y_i(\mathbf{t}_i) | Z_i(\mathbf{t}_i), \sigma_\epsilon^2) f(\sigma_\epsilon^2).$$

Step 5: Given current value of $\Sigma_\tau = \Sigma_Z(\boldsymbol{\tau}, \boldsymbol{\tau})$, we can update the hierarchical parameter σ_s^2 from

$$\sigma_s^2 | \Sigma_\tau \sim G\left(a_s + \frac{(\delta + K - 1)K}{2}, b_s + \frac{1}{2} \text{trace}(\mathbf{A}(\boldsymbol{\tau}, \boldsymbol{\tau})\Sigma_\tau^{-1})\right), \quad (13)$$

which is derived from

$$f(\sigma_s^2 | \Sigma_\tau) \propto f(\Sigma_\tau | \sigma_s^2) f(\sigma_s^2).$$

With a fairly large number of MCMC iterations (e.g., 10,000 in our numerical studies), the posterior samples will generally pass the convergence diagnosis by potential scale reduction factor (PSRF) (Gelman and Rubin, 1992).

4. Simulation studies

By simulations, we compared the performance of BABF to some of the existing smoothing and estimation methods for functional data: cubic smoothing spline (CSS) by Green and Silverman (1993), PACE by Yao et al. (2005a), Bayesian functional principle component analysis (BFPCA) by Crainiceanu and Goldsmith (2010), standard Bayesian GP regression (BGP) by Gibbs (1998), and previous BHM method (denoted by BHM in the following context) by Yang et al. (2015). Comparisons were conducted for both stationary and nonstationary functional data, both common observation grids and random observation grids (an extreme case of uncommon observation grids).

Note that both BFPCA and BGP were developed with common grids; BHM has computational issues with extremely large pooled-grid (the case with random grids); and BHM is known to be comparable with PACE (Yang et al., 2015). Thus, we compared all methods in the simulation studies with common grids, but only compared BABF to CSS and PACE in the simulation studies with random grids.

In the simulation studies, we applied CSS to each simulated functional curve independently with smoothing parameter selected by general cross-validation (GCV). For BFPCA, the covariance estimate by PACE was used, and the number of principle functions was selected subject to reserving 99.99% data variance. For the covariance function in BGP, we assumed the Matérn model with stationary data, while fixed the covariance structure at the PACE estimate with nonstationary data.

All MCMC samples consisted of 2,000 burn-ins and 10,000 posterior samples, and passed the convergence diagnosis by PSRF (Gelman and Rubin, 1992).

4.1 Studies with common grids

We generated 30 stationary functional curves (true signals) on the common equally-spaced-grid with length 40, over $\mathcal{T} = (0, \pi/2)$, from

$$GP(\mu(t) = 3 \sin(4t), \Sigma(s, t) = 5 \text{Matern}_{cor}(|s - t|; \rho = 0.5, \nu = 3.5)), \quad (14)$$

denoted by \mathbf{Z} . Specifically,

$$\text{Matern}_{cor}(d; \rho, \nu) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu} \frac{d}{\rho} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{d}{\rho} \right), d \geq 0, \rho > 0, \nu > 0, \quad (15)$$

where ρ is the scale parameter; ν is the order of smoothness; $\Gamma(\cdot)$ is the gamma function and $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind. The noise terms $\{\epsilon_{ij}\}$ were generated i.i.d. from $N(0, \sigma_{\epsilon} = \sqrt{5}/2)$, such that the signal to noise ratio (SNR) is 2 (resulting relatively high volume of noise in the simulated data). The observed noisy functional data curves were given by $\mathbf{Y} = \mathbf{Z} + \boldsymbol{\epsilon}$.

Similarly, we generated 30 nonstationary functional curves on the same equally-spaced-grid with length 40, over $\mathcal{T} = (0, \pi/2)$, from a nonstationary GP $\tilde{X}(t) = h(t)X(s(t))$ (i.e., a nonlinear

transformation of a stationary GP $X(\cdot)$; where $X(\cdot)$ denotes the GP (14), $h(t) = t + 1/2$, $s(t) = t^{2/3}$. Noisy observation data were obtained by adding noises from $N(0, \sigma_\epsilon = \sqrt{5}/2)$ to the generated nonstationary GP data (true signals).

Both stationary and nonstationary functional datasets were simulated for 100 times, on which we applied BABF (with an equally spaced working grid $\tau_{1 \times 20} \subset \mathcal{T}$) and other five alternative methods. We calculated the average root mean square errors (RMSEs) for estimates of signals $\{Z_i(t)\}$, mean function $\mu_Z(t)$, covariance surface $\Sigma_Z(t, t)$, and noise variance σ_ϵ^2 (t is the common grid). The average RMSEs (with standard deviations among 100 repeated simulations) for stationary data are shown in Table 1, while the ones for nonstationary data are shown in Table 2. Average RMSEs are omitted if the parameters are not directly estimated by the corresponding methods. For examples, BFPCA dose not directly estimate μ_Z , Σ_Z , and σ_ϵ^2 ; CSS dose not directly estimate σ_ϵ^2 . The CSS estimates of μ_Z and Σ_Z are given by the sample estimates using the pre-smoothed signals by CSS.

[Table 1 about here.]

From Table 1, we can see that, in the case with stationary data and common grids, BGP gave the best estimates for signals and noise variance (lowest RMSEs); while BHM and BABF gave the second best estimates for signals and noise variance, but the best mean-covariance estimates. From Table 2, we can see that, in the case with nonstationary data and common grids, BGP and PACE produced the best covariance estimates; while BABF produced closely accurate covariance estimates, but the best estimates for signals, mean function, and noise variance. Due to more stable computations with nonstationary data, our BABF method produced better estimates than BHM. In addition, the CSS and BFPCA methods produced least accurate estimates (with the highest RMSEs) for both stationary and nonstationary data (Tables 1-2), which demonstrates the advantages of simultaneously smoothing and estimating functional data as in BGP, BHM, and BABF.

[Table 2 about here.]

[Figure 1 about here.]

In Figure 1 (a, b), we present two sample functional observations (circles) from the stationary and nonstationary datasets, along with the underlying true signals (solid lines), smooth estimates by BHM (dash-dot lines), BGP (dotted lines), and BABF (dashed lines). Mean signal estimates by BHM (dash-dot lines), BGP (dotted lines), and BABF (dashed lines) are plotted in Figure 1 (c, d), along with the sample mean estimates (circles) and true mean curves (solid lines). The 95% pointwise credible intervals (crosses) for the signal estimates and mean estimate by BABF are also added to the plots in Figure 1. We can see that, with common grids, all three Bayesian methods produced similarly good estimates for functional signals and mean. Especially, with stationary data, the estimates for functional signals are almost overlapped with each other. However, with nonstationary data, our BABF method produced the best smooth estimates for functional signals.

We also present 3 dimensional (3D) plots for the covariance surface estimates by BGP, BHM and BABF in Figure 2 (a, b, c, e, f, g), along with the true covariance surfaces in Figure 2 (d, h) for the stationary data (with a Matérn surface) and nonstationary data (with a nonlinearly transformed Matérn surface). Because BGP assumes a parametric Matérn model, the true covariance model for simulated stationary data, the functional signals and mean estimates by BGP have relatively low RMSEs. Although the covariance estimate by BGP is still a Matérn function with the most accurate structure, the variances were underestimated, compared to BHM and BABF (Figure 2). However, in real data analysis, a wrong model is usually assumed for the covariance surface in BGP, which is likely to produce estimates with large errors. In contrast, the nonparametric covariance estimates by BHM and BABF would be more flexible and applicable in real data analysis.

[Figure 2 about here.]

In addition, we present the coverage probabilities of the 95% pointwise credible intervals by BGP, BHM, and BABF in one example simulation, as in Table 3. For functional signals, BGP has the highest coverage probability with stationary data, while BHM and BABF have higher

coverage probabilities with nonstationary data. All methods have similar coverage probabilities for the functional mean, where the relatively low values are due to narrow 95% confidence intervals. As for the covariance function, the coverage probability of BGP is significantly lower than the ones of BHM and BABF for both stationary and nonstationary data, because of the variance underestimation by BGP.

[Table 3 about here.]

In summary, with common grids, Bayesian GP based regression methods (BGP, BHM, and BABF) produced better smoothing and estimation results, compared to estimating mean-covariance functions with pre-smoothed functional data by CSS. BABF gave at least similar results (better results with nonstationary data) as the previous BHM.

4.2 Studies with random grids

For each simulated nonstationary dataset, we generated 30 true functional curves from the stationary and non-stationary GPs as in Section 4.1, with observational grids (length 40) that are randomly (uniformly) generated over $\mathcal{T} = (0, \pi/2)$. Raw functional data were then obtained by adding noises from $N(0, \sqrt{5}/2)$ to the true functional data. We compared our BABF method (with an equally spaced working grid $\tau_{1 \times 20} \subset \mathcal{T}$) to CSS and PACE, with 100 simulation replicates.

In Table 4, we presented the average RMSEs for the estimates of signals, noise variance, mean-covariance functions (evaluated on the grid that is equally spaced over \mathcal{T} with length 40), along with standard errors of these average RMSEs in the parentheses. We can see that, for both stationary and nonstationary data with random grids, our BABF method performs consistently better than CSS and PACE (with lowest RMSEs), for both signal and mean estimation.

[Table 4 about here.]

We presented two sample observations (circles) from the stationary and nonstationary functional data in Figure 3 (a, b), along with signal estimates by CSS (dash-dot lines), PACE (dotted lines),

BABF (dashed lines), and the true signals (solid lines). We can see that, with random grids, BABF produced the best signal estimates. This is because CSS smoothed each functional curve independently, PACE only used limited information per pooled-grid point in the random-grid case, and BABF used more information through basis function approximations. For both stationary and nonstationary functional data, PACE and BABF gave closely good mean estimates, but CSS gave the worst mean estimate, as shown in Figure 3 (c, d). The 95% pointwise credible intervals (crosses) by BABF are added to the plots in Figure 3.

[Figure 3 about here.]

[Figure 4 about here.]

Covariance surface estimates by CSS, PACE and BABF (for stationary and nonstationary functional data) are plotted in Figure 4 (a, b, c, e, f, g), along with the true stationary and nonstationary covariances in 4(d, h). We can see that PACE produced the most roughest covariance estimate, which is due to using limited information on the pooled-grid points. Coverage probabilities for the BABF estimates of one example simulation are listed in Table 5, which shows that our BABF method performs very well with both stationary and nonstationary functional data. The relatively low coverage probabilities for nonstationary signals is due to narrow 95% confidence intervals.

[Table 5 about here.]

In summary, with random grids, our BABF methods produced the best signal and mean estimates, compared to CSS and PACE. Although the sample covariance estimate using pre-smoothed data by CSS has the lowest RMSE with nonstationary data, the sample covariance estimate using the smoothed data by BABF will have at least similar RMSE because of more accurately smoothed functional data. Thus, our approach of conducting Bayesian references with basis function approximations is indeed valid, with the advantages of efficient and stable computations.

5. Application on real data

We analyzed a functional dataset from a study of obesity in children and adolescents (Lee et al., 2016), by the Children's Nutrition Research Center (CNRC) at Baylor College of Medicine. This study estimated the energy expenditure (EE, minute by minute, in unit kcal) of 106 children and adolescents (44 obese cases, 62 nonobese controls) during sleep, by using the CNRC room respiration calorimeters (Moon et al., 1995). The original data were collected during 24 hours with a series of scheduled physical activities for participants, however, we only used the sleeping EE (SEE) data measured at 405 time points during the sleeping period (12:00am-7:00am), where each participant has a SEE signal curve (a functional observation over time) with 405 observations. This real SEE data set is a good example of the case with high dimensional common grids. The goal of this study is to identify different EE patterns between obese cases and controls, and then provide insights about using EE to make obesity predictions.

Many statistical tools have been used to analyze this type of functional data. For example, Zakeri et al. (2010) analyzed a similar EE data set by multivariate adaptive regression splines; Lee et al. (2016) analyzed the same SEE data by first smoothing the functional data by the smoothing spline method, and then using the functional principle component analysis (FPCA) method to identify patterns in the obese and nonobese groups. Here, we applied CSS, PACE, and our BABF method on this SEE functional dataset. Specifically, CSS was applied independently on each signal with a smoothing parameter chosen by GCV, PACE was applied with common grid $1 : 405$, and BABF was applied with the equally-spaced-grid over $[1, 405]$ with length 30 as the working grid τ . Note that both PACE and BABF were applied separately for the functional data of obese and nonobese groups.

[Figure 5 about here.]

We present one example signal (circles) from each group, along with signal estimates by CSS (dash-dot lines), PACE (dotted lines), and BABF (dashed lines) in Figure 5 (a, b). The estimates

of group means are plotted in Figure 5 (c, d), including estimates by the sample mean (circles), CSS (dash-dot lines), PACE (dotted lines), and BABF (dash-dot lines). Pointwise 95% credible intervals (crosses) by BABF are added to the plots. Figures 6 and 7 show 3D plots of the covariance surface estimates and heatmaps of the correlation estimates by CSS, PACE, and BABF. Again, the mean-covariance estimates given by CSS are the sample estimates with the pre-smoothed signals.

Clearly, CSS produced the roughest signal estimates, which lead to the roughest mean-covariance estimates (Figures 5-7). Both PACE and our BABF method gave reasonable smooth signal estimates (Figure 5 (a, b)) and covariance estimates (Figure 6 (b, c, e, f)), showing the periodic patterns of the SEE data. Note that PACE is essentially a FPCA method with smoothed mean-covariance estimates, which is expected to give similar results as in Lee et al. (2016). Whereas, the mean estimates by our BABF method show better periodic patterns than PACE (Figure 5 (c, d)), and the correlation estimates by BABF show less correlations between two apart time points than the ones by PACE (Figure 7 (b, c, e, f)). This real case study demonstrates that our BABF method is a good statistical tool for functional data analysis.

[Figure 6 about here.]

[Figure 7 about here.]

6. Discussion

In this paper, we propose a novel approach that conducts posterior inferences for Bayesian GP regression models (e.g., BHM by Yang et al. (2015)) with basis function approximations for functional data analysis. Our approach represents functional data by a system of basis functions, and then conducts posterior references through MCMC samplings of the basis coefficients. Specifically, we show details about deriving the induced Bayesian framework and MCMC process for the basis coefficients, with the BHM. As a result, the BHM with approximations by basis functions (BABF) not only retains the same advantages as BHM, such

as simultaneously smoothing (without the necessity of selecting smoothing parameters) and estimating mean-covariance functions in a nonparametric way; but also provides one solution for the computational bottleneck of the Bayesian GP regression methods. In addition, by MCMC samplings with the basis coefficients that usually have a much smaller dimension K , compared to the pooled-grid dimension p , the BABF reduces the computation complexity from $O(np^3m)$ to $O(nK^3m)$ with sample size n and total MCMC iterations m .

Both simulation and real case studies demonstrate that BABF has similar performance as BHM and other Bayesian GP regression models in the case with low dimensional common grids, and that BABF outperforms the alternative methods (e.g., CSS, PACE) in the case with random grids and the case with high dimensional common grids. Although the computation time by BABF (with 12,000 MCMC iterations) is still about 4 times of the computation time by the frequentist method PACE, BABF provides complementary credible intervals to quantify the uncertainties of parameter estimation, and basis function representations for the nonparametric estimates of functional signals, mean, and covariance.

Of course, good performance of our method with basis function approximations is related to a set of reasonable basis functions. The type of selected basis functions should depend on the analyzed data, e.g., Fourier series could be used for periodic data, B-splines could be used for Gaussian process data (as used in this paper), and wavelets could be used for signal data. The general strategies of choosing basis functions for representing data curves are applied to our method, which is not sensitive to the number of basis function, as long as a reasonable number is used (e.g., 20 in our simulations, 30 in the study of SEE dataset). Take B-splines for an example, the knots used for constructing B-splines should depend on the density of the analyzed data over the domain. We used the `optknt` function in Matlab to obtain a knot sequence that is best for interpretation at the working grid τ (Gaffney and Powell, 1976; Micchelli et al., 1976; De Boor, 1977), where τ should be chosen to well represent data densities over the domain (e.g., equally-spaced for our simulated

data that have either equally-spaced observation grids or uniformly generated random grids). More details about constructing appropriate B-splines can be found in De Boor et al. (1978).

In conclusion, our BABF method sheds light on efficiently conducting Bayesian functional data analysis with Gaussian-Wishart processes and interpreting covariance functions by basis functions. A software for implementing the BHM and BABF methods is freely available at <https://github.com/yjingj/BFDA>.

Acknowledgments

The authors would like to thank the Children’s Nutrition Research Center at the Baylor College of Medicine for providing the metabolic data (funded by National Institute of Diabetes and Digestive and Kidney Diseases Grant DK-74387 and the USDA/ARS under Cooperative Agreement 6250-51000-037). Jingjing Yang and Dennis D. Cox were supported by the NIH grant PO1-CA-082710.

References

- Abraham, C., Cornillon, P.-A., Matzner-Løber, E., and Molinari, N. (2003). Unsupervised curve clustering using b-splines. *Scandinavian Journal of Statistics* **30**, 581–595.
- Banerjee, A., Dunson, D. B., and Tokdar, S. T. (2013). Efficient Gaussian process regression for large datasets. *Biometrika* **100**, 75–89.
- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 825–848.
- Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45**, 11–22.
- Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13**, 571–592.

- Crainiceanu, C. M. and Goldsmith, A. J. (2010). Bayesian functional data analysis using winbugs. *Journal of Statistical Software* **32**, 11.
- Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 209–226.
- Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a bayesian application. *Biometrika* **68**, 265–274.
- De Boor, C. (1977). *Computational aspects of optimal recovery*. Springer US, Boston, MA.
- De Boor, C., De Boor, C., De Boor, C., and De Boor, C. (1978). *A practical guide to splines*, volume 27. Springer-Verlag New York.
- Gaffney, P. W. and Powell, M. J. D. (1976). *Optimal interpolation*. Springer Berlin Heidelberg.
- Garcia-Escudero, L. A. and Gordaliza, A. (2005). A proposal for robust curve clustering. *Journal of classification* **22**, 185–201.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* pages 457–472.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **6**, 721–741.
- Gibbs, M. N. (1998). *Bayesian Gaussian processes for regression and classification*. PhD thesis, University of Cambridge, UK.
- Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press.
- Hall, P., Horowitz, J. L., et al. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35**, 70–91.
- James, M. (1978). The generalised inverse. *The Mathematical Gazette* pages 109–114.
- Lee, J. S., Zakeri, I. F., and Butte, N. F. (2016). Functional principal component analysis and

- classification methods applied to dynamic energy expenditure measurements in children. *Technical Report* .
- Matérn, B. (1960). *Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations*. PhD thesis, Meddelanden fran Statens Skogsforskningsinstitut.
- Micchelli, C. A., Rivlin, T. J., and Winograd, S. (1976). The optimal recovery of smooth functions. *Numerische Mathematik* **26**, 191–200.
- Moon, J. K., Vohra, F. A., Valerio Jimenez, O. S., Puyau, M. R., and Butte, N. F. (1995). Closed-loop control of carbon dioxide concentration and pressure improves response of room respiration calorimeters. *Journal of Nutrition-Baltimore and Springfield then Bethesda-* **125**, 220–220.
- Quiñonero Candela, J., E., R. C., and Williams, C. K. I. (2007). Approximation methods for gaussian process regression. Technical report, Applied Games, Microsoft Research Ltd.
- Ramsay, J. O. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 539–572.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA.
- Serban, N. and Wasserman, L. (2012). Cats: Clustering after transformation and smoothing. *Journal of the American Statistical Association* .
- Shi, J., Wang, B., Murray-Smith, R., and Titterton, D. (2007). Gaussian process functional regression modeling for batch data. *Biometrics* **63**, 714–723.
- Shi, J. Q. and Choi, T. (2011). *Gaussian process regression analysis for functional data*. CRC Press, Boca Raton, FL.

- Yang, J., Zhu, H., Choi, T., and Cox, D. D. (2015). Smoothing and mean-covariance estimation of functional data with a bayesian hierarchical model. *Bayesian Analysis* .
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33**, 2873–2903.
- Zakeri, I. F., Adolph, A. L., Puyau, M. R., Vohra, F. A., and Butte, N. F. (2010). Multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents. *Journal of Applied Physiology* **108**, 128–136.
- Zhu, H., Brown, P. J., and Morris, J. S. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association* **106**,.
- Zhu, H. and Cox, D. D. (2009). A functional generalized linear model with curve selection in cervical pre-cancer diagnosis using fluorescence spectroscopy. In *Optimality: The Third Erich L. Lehmann Symposium*, volume 57 of *Lecture Notes Monograph Series*, pages 173–189. Institute of Mathematical Statistics, Beachwood, OH, USA.
- Zhu, H., Vannucci, M., and Cox, D. D. (2010). A Bayesian hierarchical model for classification with selection of functional predictors. *Biometrics* **66**, 463–473.
- Zhu, H., Yao, F., and Zhang, H. H. (2014). Structured functional additive regression in reproducing kernel hilbert spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 581–603.

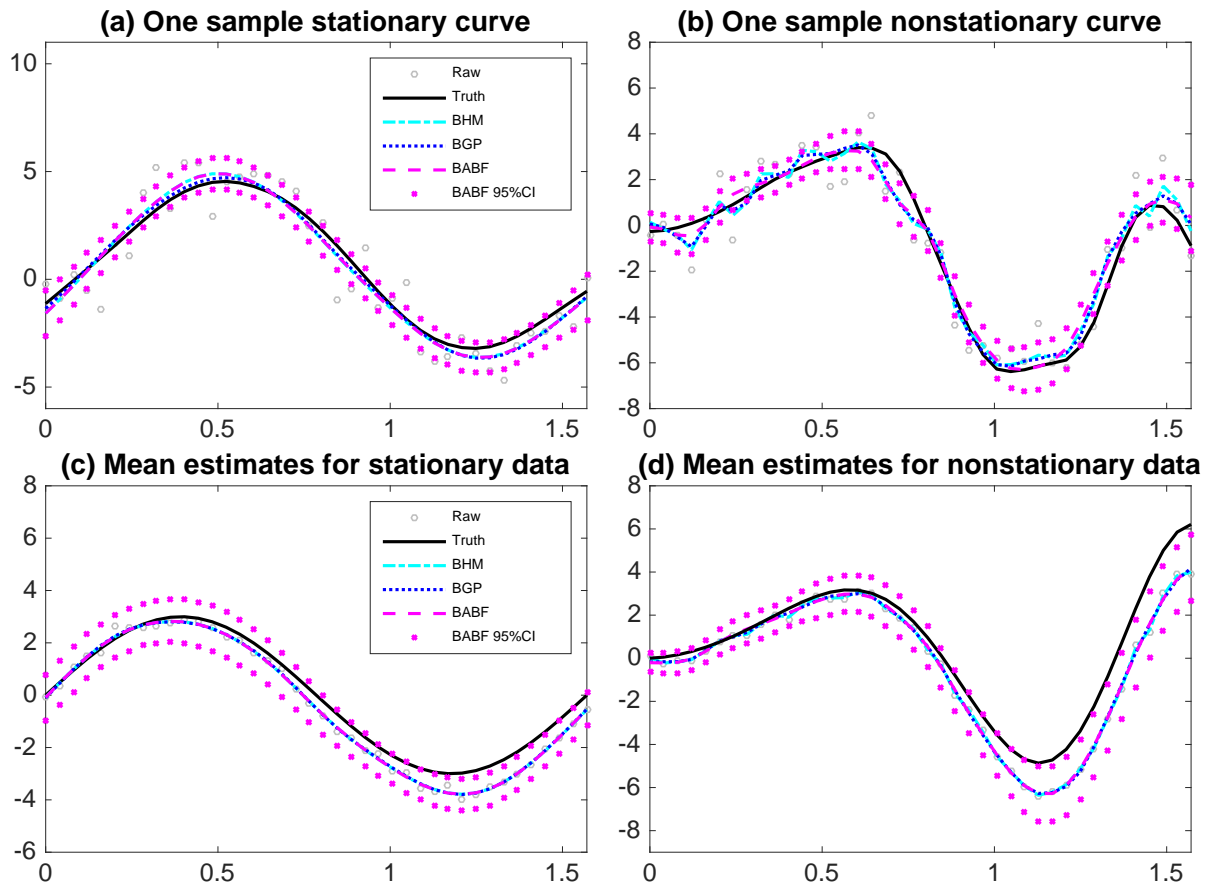


Figure 1. Sample curve estimates in (a), (b); mean estimates in (c), (d).

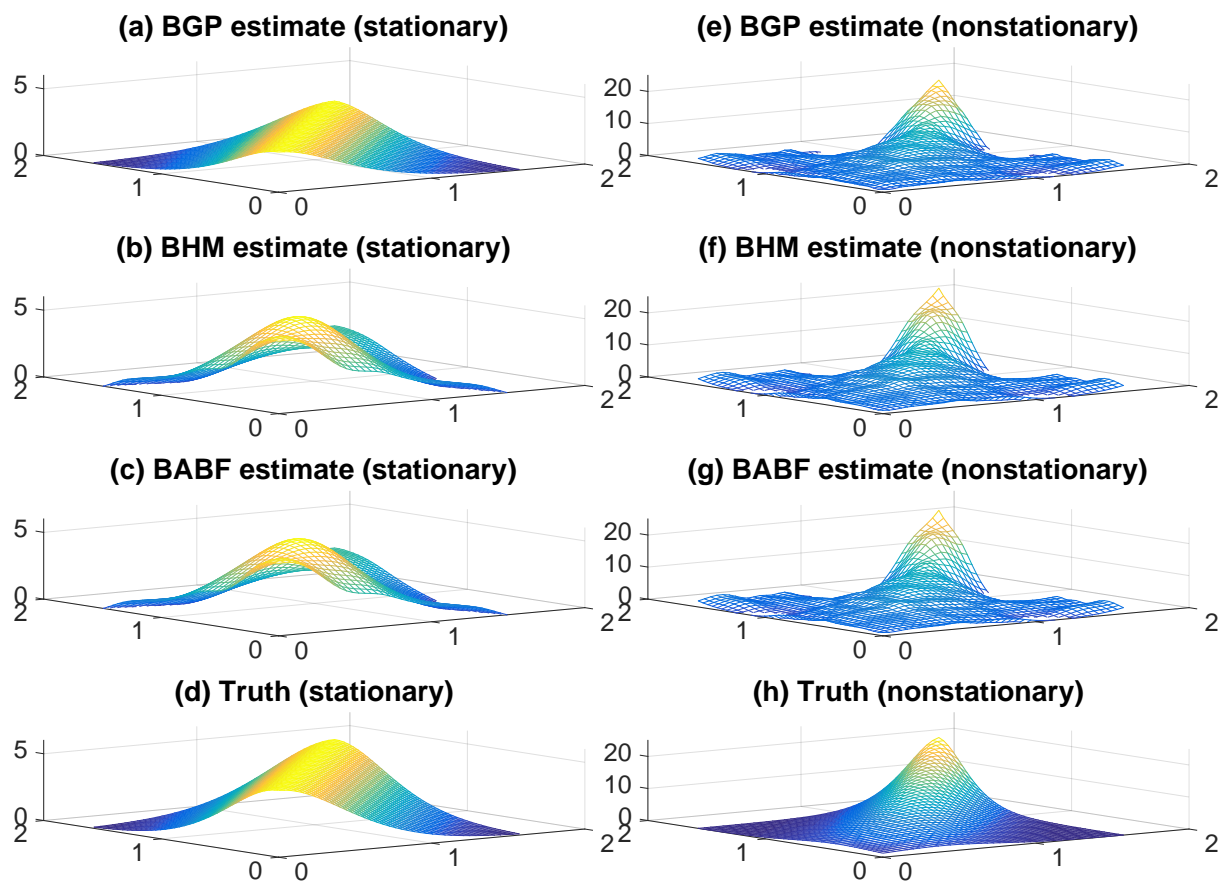


Figure 2. Covariance estimates and the true covariances.

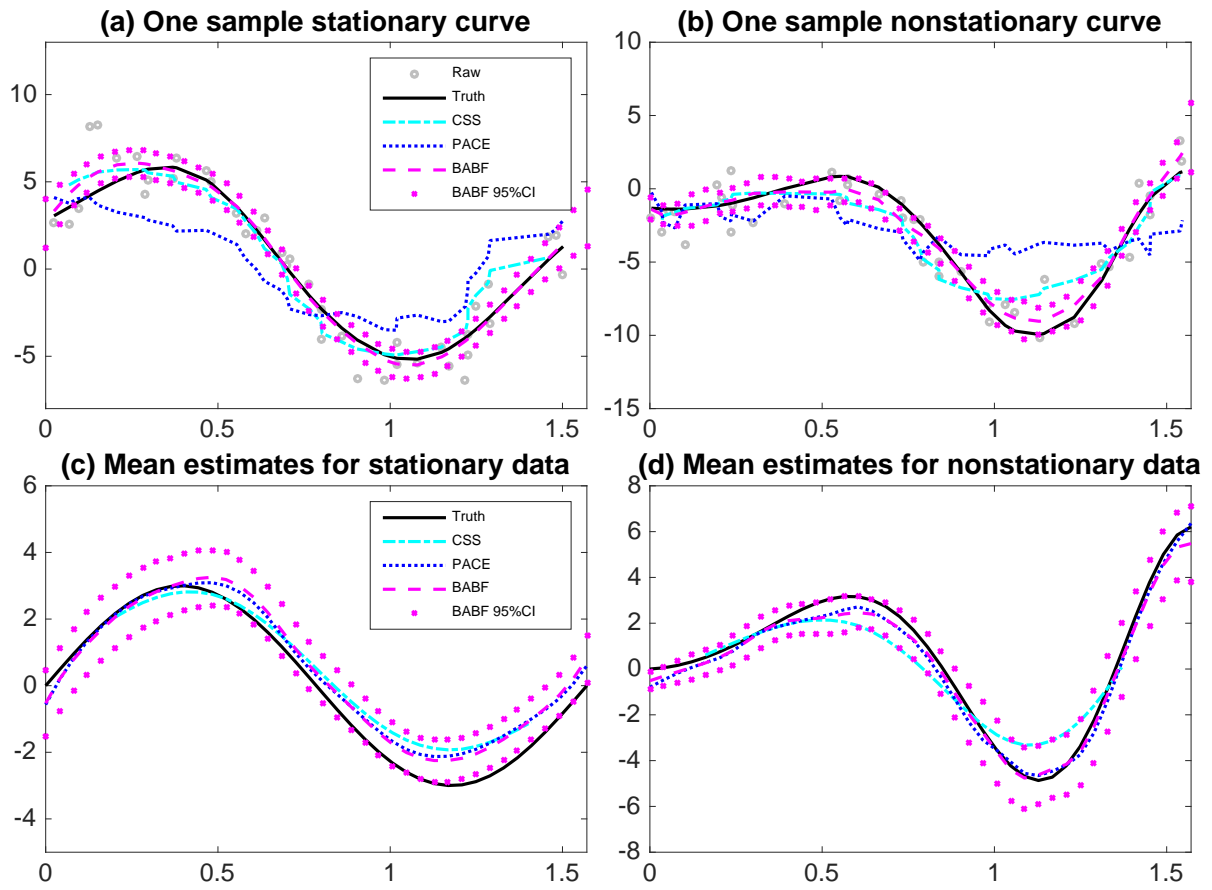


Figure 3. Two sample curve estimates in (a), (b); and mean estimates in (c), (d).

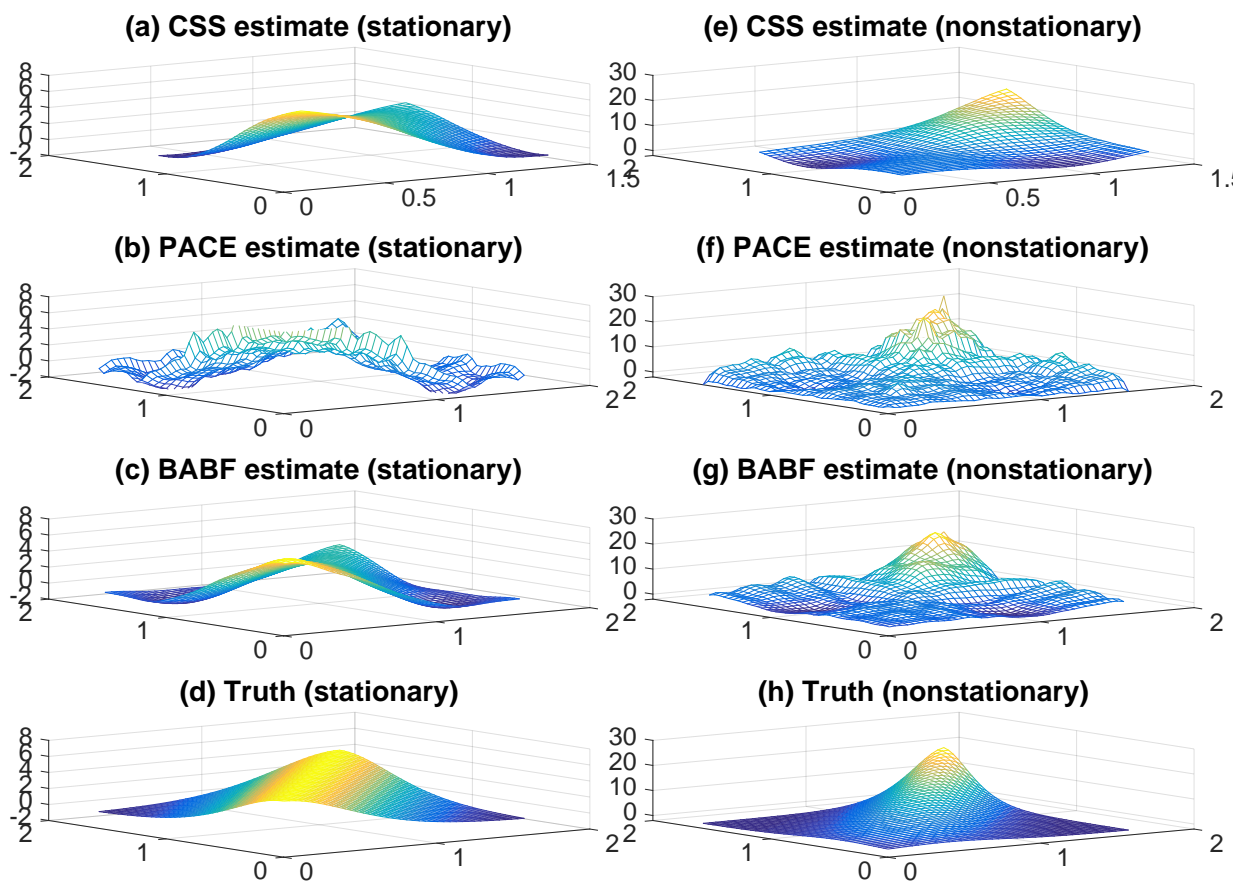


Figure 4. Covariance estimates and the true covariances.

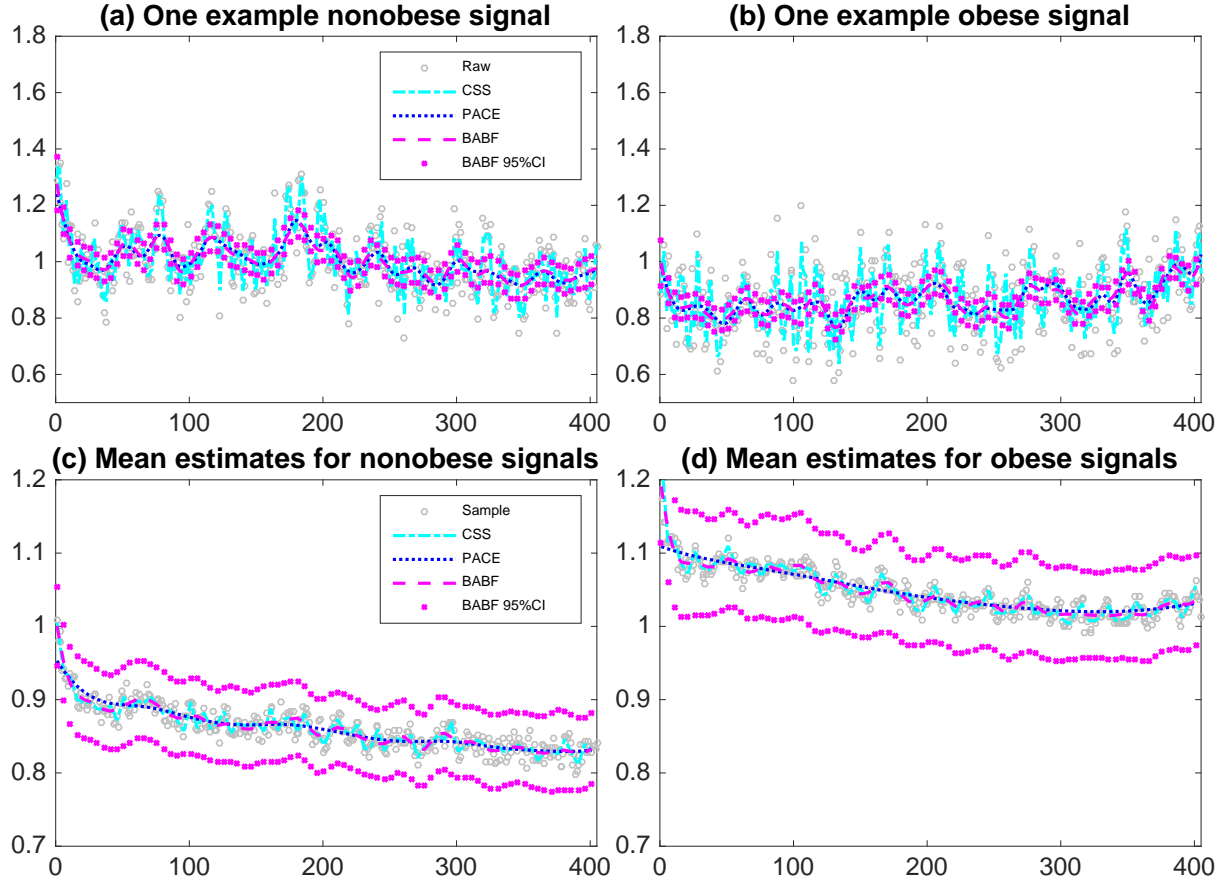


Figure 5. Example signal estimates by CSS, PACE, and BABF, along with 95% intervals by BABF in frames (a, b); mean estimates by sample mean, CSS, PACE, and BABF, along with 95% intervals by BABF in frames (c, d).

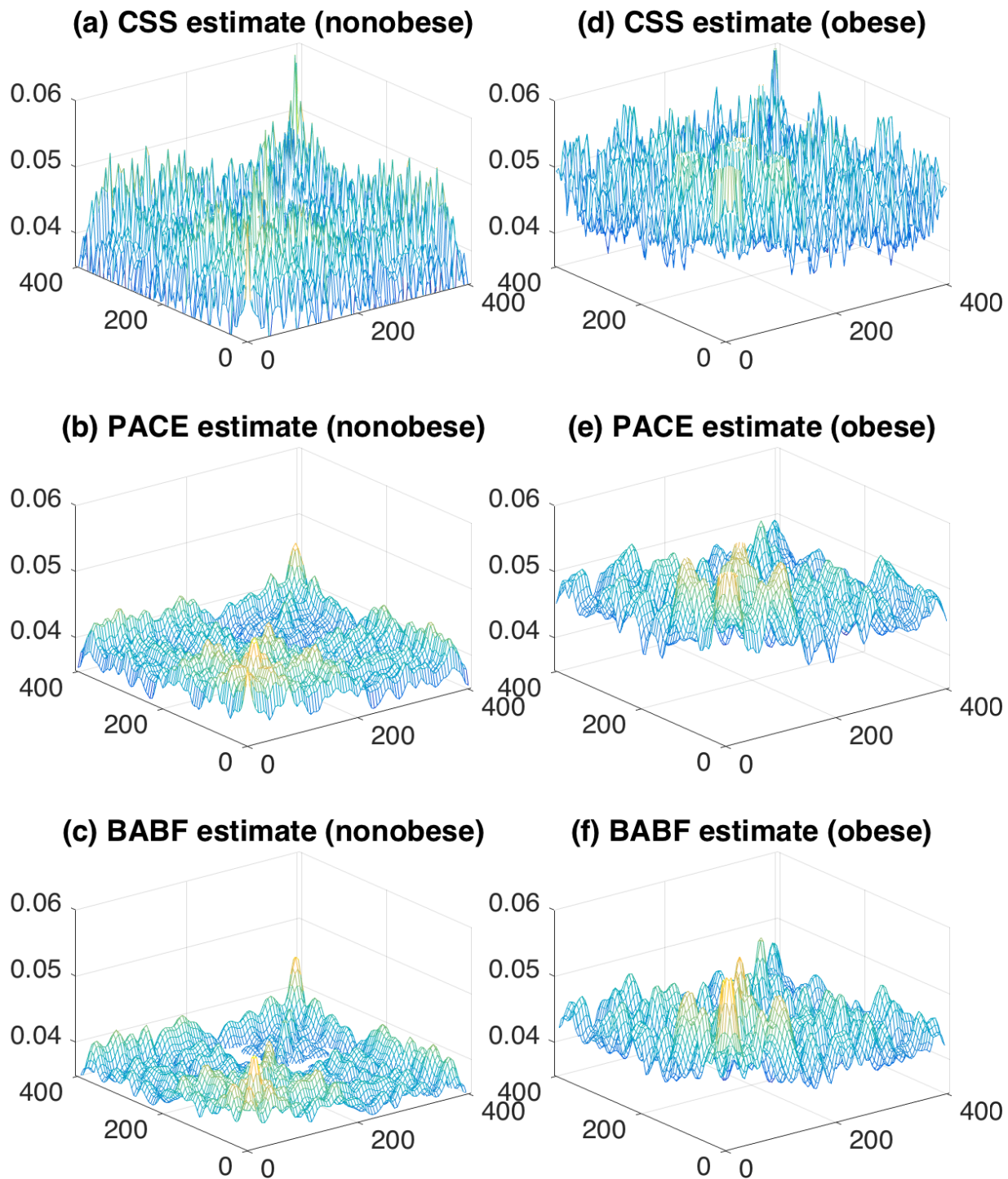


Figure 6. Covariance estimates by CSS, PACE and BABF.

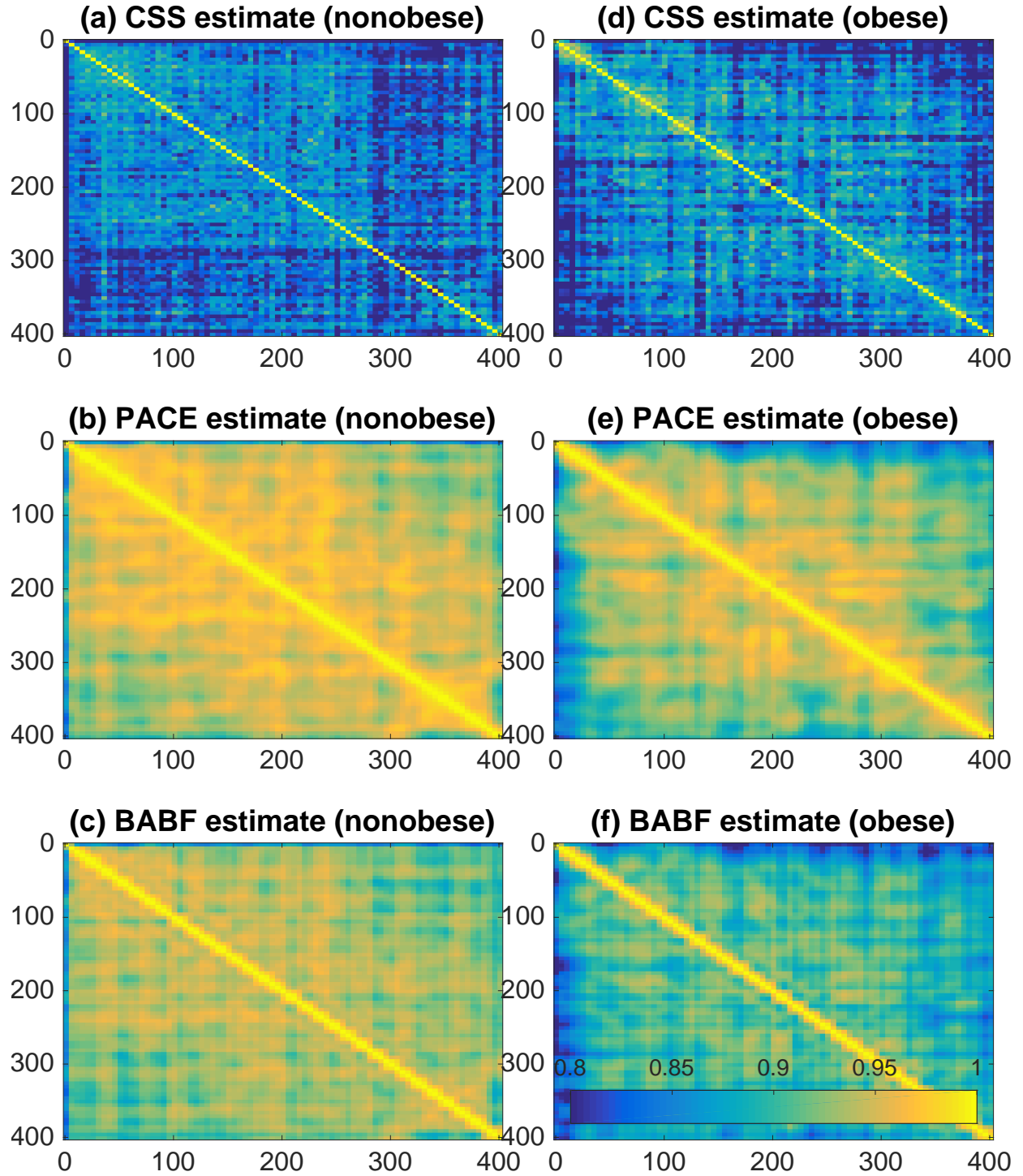


Figure 7. Heatmaps of the correlation estimates by CSS, PACE and BABF.

Table 1

Simulation results with **stationary** functional data and common grids: average RMSEs and corresponding standard errors (in parentheses) of $\{Z_i(\mathbf{t})\}$, $\mu(\mathbf{t})$, $\Sigma_Z(\mathbf{t}, \mathbf{t})$, and σ_ϵ^2 produced by CSS, PACE, BFPCA, BGP, BHM, and BABF. Average RMSEs are omitted if the corresponding parameters are not directly estimated. Two best results are bold for each parameter.

	CSS	PACE	BFPCA	BGP	BHM	BABF
$\{Z_i(\mathbf{t})\}$	0.4808 (0.0213)	0.4553 (0.0268)	0.5657 (0.0550)	0.4020 (0.0219)	0.4067 (0.0207)	0.4073 (0.0204)
$\mu(\mathbf{t})$	0.4757 (0.1347)	0.4194 (0.1593)	- -	0.3982 (0.1527)	0.3961 (0.1538)	0.3961 (0.1535)
$\Sigma(\mathbf{t}, \mathbf{t})$	1.0017 (0.3079)	1.0375 (0.2850)	- -	1.0988 (0.4934)	0.9601 (0.2902)	0.9590 (0.2913)
σ_ϵ^2	- -	0.0764 (0.0516)	- -	0.0460 (0.0327)	0.0491 (0.0357)	0.0483 (0.0352)

Table 2

Simulation results with **nonstationary** functional data and common grids: average RMSEs and corresponding standard errors (in parentheses) of $\{Z_i(\mathbf{t})\}$, $\mu(\mathbf{t})$, $\Sigma_Z(\mathbf{t}, \mathbf{t})$, and σ_ϵ^2 by CSS, PACE, BFPCA, BGP, BHM, and BABF. Average RMSEs are omitted if the corresponding parameters are not directly estimated. Two best results are bold for each parameter.

	CSS	PACE	BFPCA	BGP	BHM	BABF
$\{Z_i(\mathbf{t})\}$	1.0271 (0.00463)	0.5185 (0.0255)	0.6314 (0.0632)	0.5183 (0.0265)	0.5759 (0.0227)	0.5133 (0.0227)
$\mu(\mathbf{t})$	0.9446 (0.1509)	0.5782 (0.2095)	-	0.5387 (0.2090)	0.5530 (0.2038)	0.5356 (0.2094)
$\Sigma(\mathbf{t}, \mathbf{t})$	1.9635 (0.8386)	1.9751 (0.8160)	-	1.9733 (0.6831)	2.0296 (0.6891)	1.9768 (0.7835)
σ_ϵ^2	-	0.0810 (0.0541)	-	0.1472 (0.0879)	0.2432 (0.0644)	0.0692 (0.0492)

Table 3

Example coverage probabilities for the 95% pointwise credible intervals of $\{Z_i(\mathbf{t})\}$, $\mu(\mathbf{t})$, and $\Sigma(\mathbf{t}, \mathbf{t})$ by BGP, BHM and BABF. The highest probability is bold for each parameter.

		BGP	BHM	BABF
$\{Z_i(\mathbf{t})\}$	stationary	0.9483	0.9217	0.9208
	nonstationary	0.8350	0.9450	0.8742
$\mu(\mathbf{t})$	stationary	0.7500	0.7250	0.7250
	nonstationary	0.6750	0.6750	0.6750
$\Sigma(\mathbf{t}, \mathbf{t})$	stationary	0.0000	0.7869	0.7869
	nonstationary	0.3819	0.9913	0.9938

Table 4

Simulation results with *stationary/nonstationary* functional data and random grids: average RMSEs and corresponding standard errors (in parentheses) of $\{Z_i(\mathbf{t})\}$, $\mu(\mathbf{t})$, $\Sigma_Z(\mathbf{t}, \mathbf{t})$, and σ_ϵ^2 by CSS, PACE, and BABF. Average RMSEs are omitted if the corresponding parameters are not directly estimated. Best results are bold for each parameter.

	Stationary			Nonstationary		
	CSS	PACE	BABF	CSS	PACE	BABF
$\{Z_i(\mathbf{t})\}$	0.4839 (0.0229)	1.4141 (0.1424)	0.4079 (0.0219)	1.0137 (0.0511)	2.6300 (0.2876)	0.6832 (0.0576)
$\mu(\mathbf{t})$	0.4229 (0.1471)	0.4196 (0.1290)	0.3690 (0.1302)	0.9905 (0.1888)	0.6157 (0.2160)	0.5920 (0.2138)
$\Sigma(\mathbf{t}, \mathbf{t})$	1.0445 0.4313	1.4089 (0.3502)	1.0054 (0.3286)	1.6403 (0.6086)	2.4120 (0.6497)	2.2090 (0.4506)
σ_ϵ^2	- -	0.1900 (0.1818)	0.0509 (0.0387)	- -	0.4007 (0.2960)	0.2209 (0.1189)

Table 5

Example coverage probabilities for the 95% pointwise credible intervals of $\{Z_i(\mathbf{t})\}$, $\mu(\mathbf{t})$, and $\Sigma(\mathbf{t}, \mathbf{t})$ by BABF.

	$\{Z_i(\mathbf{t})\}$	$\mu(\mathbf{t})$	$\Sigma(\mathbf{t}, \mathbf{t})$
stationary	0.8650	1.0000	0.9506
nonstationary	0.5625	0.9000	0.8550